Collaborative Oceanography and Virtual Experiments

James Bellingham
Monterey Bay Aquarium Research Institute
7700 Sandholdt Road, Moss Landing, CA, 95039
(831) 775-1731; (831) 775-1646; jgb@mbari.org

Award Number: N00014-09-1-0405
https://itop.org

LONG-TERM GOALS

The long term goals are develop an online collaboration environment that supports cooperative development and operation of complex environmental field programs by distributed interdisciplinary teams. We are using these tools to facilitate ITOP virtual experiments and the ITOP field program, and packaging the collaboration tools for re-use by ONR and other entities.

OBJECTIVES

The primary technological objectives of this effort are to provide core tools to facilitate data aggregation and discussion, to provide tools that integrate data and analysis with discussions, to provide graphical overviews that show linkages between discussions and intersections of data with discussions, and to provide these tools in a reusable package that can be easily reused in future ONR field experiments and observatories. This combination addresses a unique need of scientific collaboration, and will greatly improve the ability of investigators to rapidly understand not just data, but also the discussions and decisions based upon that data.

APPROACH

Previous work created and demonstrated a data collection and collaboration system tailored for the Adaptive Sampling and Prediction (ASAP) component of the Monterey Bay 2006 Experiment[i]. This work is focusing on advancing the highly-tailored tools developed for prior field programs[ii], specifically the COOP[iii] (Collaborative Ocean Observatory Portal) into a comprehensive set of elements applicable to the ITOP and other follow-on programs.

Development is focusing on three interdependent categories: 1) data system tools such as data collection mechanisms, data conversion, data exploration, and visualization; 2) human collaboration tools, such as discussion forums and data annotation, both of which will be provided via a web-accessible portal; and 3) portal design issues that will affect how the portal can be adopted and re-used by future field programs.

Report Documentation Page						*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.						
1. REPORT DATE **2010**		2. REPORT TYPE				3. DATES COVERED **00-00-2010 to 00-00-2010**
4. TITLE AND SUBTITLE **Collaborative Oceanography and Virtual Experiments**						5a. CONTRACT NUMBER
						5b. GRANT NUMBER
						5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)						5d. PROJECT NUMBER
						5e. TASK NUMBER
						5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Monterey Bay Aquarium Research Institute,7700 Sandholdt Road,Moss Landing,CA,95039**						8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)						10. SPONSOR/MONITOR'S ACRONYM(S)
						11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT						
15. SUBJECT TERMS						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**		

Data Systems

Development of data collection and storage tools were given the first priority, as preliminary discussions with ITOP investigators indicated a strong desire for a system that had the capabilities of NCAR's Earth Observing Laboratory (EOL), which is primarily a system for aggregating and presenting externally generated products. However, ITOP researchers also wanted to have binary field data and model outputs in a place where the program investigators can easily manipulate the data.

After binary data is collected, it must be either converted to a common format or otherwise handled via a common mechanism that provides consistent metadata. In the 2003 and 2006 AOSN field programs data was converted to a common format -- this approach was found to be reliable, but doubled data storage requirements and complicated versioning and management of data. The approach for ITOP will be to leverage UCAR's NetCDF library to "wrap" raw binary files so they look like CF-Metadata compliant NetCDF files, which can be manipulated via off-the-shelf tools such as UCAR's Integrated Data Viewer[iv].

Several lessons were learned in the Spring 2010 Virtual Experiment.

- Four-panel plots were very useful for comparing products from different sources, and generating them automatically on the server was much more efficient than manually stitching together discrete plots in an image editing program. However, there was a demand for any user to be able to modify the selection of the four panels, which was implemented for the field program.
- There was no reason to duplicate products that UCAR was collecting and managing on the EOL site, but it would be worthwhile to index those products if value can be added by the CoSci user interface (such as kml wrappers for viewing the products in Google Earth). Hence, "federated" products that remain at their origination were added to CoSci.
- There was a strong desire to view the position of ITOP ocean assets as individual kml elements that can be viewed in Google Earth, so the scripting tools built into CoSci were used to generate the kml tracks.

An active area of AOSN research, funded by both ONR and the Packard Foundation, is the data discover and data access tool called the Metadata Oriented Query Assistant (MOQuA)[v]. Our approach for FY2011 is to incorporate MOQuA features into the ITOP collaborative tool suite.

Collaborative Tools

Our approach to designing the human collaboration systems began by examining how ITOP investigators made operational decisions in the ITOP pilot experiment and virtual experiment. We learned that the decision-making process in ITOP is primarily top-down, coming from a small cadre of co-located researchers interacting verbally, rather than the asynchronous proposal and voting based process used in the ASAP experiment. A strong preference was shown for "live" web meetings, rather than the asynchronous "discussion forum" style communications that characterized the ASAP experiment.

The difference between ITOP and ASAP approaches makes sense, as ITOP is explicitly designed to be a two month experiment, while ASAP was designed to be a prototype for more permanent observation systems. It is reasonable that researchers can be co-located and operate synchronously for a few

months, but as NASA clearly learned in the Mars Exploration Rover experiment, co-location and synchronized decision making is not sustainable in the long run.

It became apparent that the tools we would be providing would be *decision influencing* tools, which would help far-flung researchers impact the centralized decision making process, rather than *decision support* tools, which would be used to garner consensus decisions.

A major lesson learned in the ASAP experiment was that investigators are often located in places (ships, for example) where they cannot interact with an on-line portal. Many investigators are also more comfortable with e-mail than on-line portals. Therefore, our approach allows investigators to contribute to discussions via e-mail, and receive updates and alerts via e-mail.

In FY2011, development will focus on an annotation system that allows investigators to tag data with comments, so that others that explore the same data collections can learn from the experience of others. Eric D'Asaro used image editing software to manually provide annotations to products during the ITOP field program in September 2010 (See Figure 1 for an example). The capability to generate these annotations through the web tool was explored in 2006 as part of COOP. This code will be updated and introduced into CoSci in 2011.

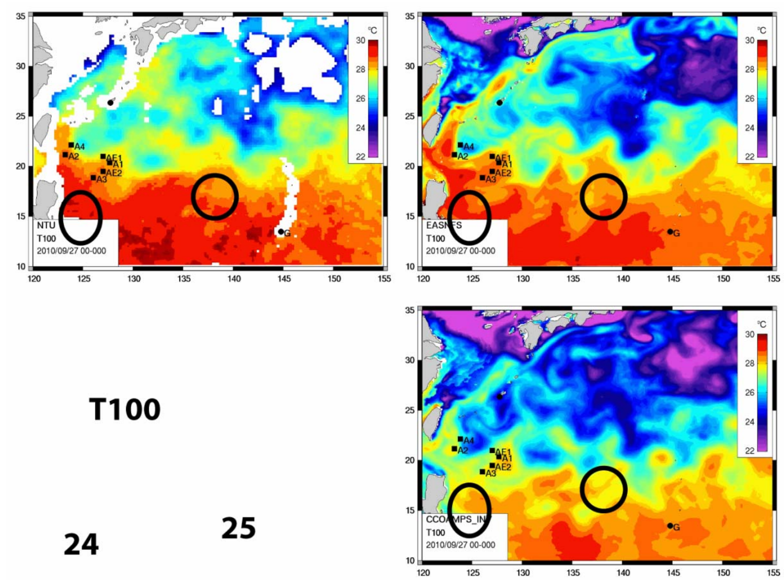

Figure 1, Multi-panel colored image created by CoSci, black annotations manually added by Eric D'Asaro using an image editing program. Goal in 2011 will be to enable such annotations inside the CoSci tool.

<u>Portal Design</u>

A critical component of the success of this program will be whether the tool is adopted and used by other experimental programs. To achieve this success, barriers to adoption must be lowered. Major emphasis in 2011 will be focused upon:

- Modularizing CoSci, so adopters can choose to deploy just those components that are critical to their investigators needs.
- Enhancing documentation built into CoSci that can assist power users in configuring data sources and installing scripts for processing data.
- Migrating CoSci module development to a standard software development model (such as JSR-286[vi]) so that developers can easily develop new features and incorporate them into CoSci, or simply import one of many already-developed portlets from vendors such as SyncEx (see http://www.syncex.com/portlet_catalog/portlet_catalog.htm) or JBoss (see http://jboss.org/portletswap/portlet_catalog.html).
- Generating installation scripts that check for and resolve missing dependencies on target servers.

WORK COMPLETED

The first versions of the Collaborative Science portal (CoSci) were released in September 2009 at https://itop.org. As requested by ITOP PIs, the system duplicated many of the features of NCAR's Earth Observing Laboratory (EOL), but also contains improvements over EOL, such as kml wrappers for any geo-referenced products and an internal architecture which will allow it to evolve into a collaborative communication tool. During September 2009, the prototype system collected over 100,000 model and observational products from repositories around the world.

In the spring of 2010, ITOP PI's used the next version of CoSci in the ITOP Virtual Experiment, which re-played the events of October 2009, but included decision making about how assets would be allocated if events were occurring in real time. The system once again collected thousands of model and observational products. However, 4-dimensional model data runs were also collected and these were used to generate the first "common" displays of ITOP ocean data. In the discussions that followed, the general formats for ocean data displays were established.

In the 2006 field experiment, processing of raw data into visual data products was carried out on a variety of machines that ran independently of the central server. While the approach worked well much of the time, it was difficult for any one "power user" to repair errors in a plotting process if they did not have read/write access to the plotting code. For ITOP, plotting was centralized on the server, and an API for loading and managing plotting scripts was developed (see Figure 2).

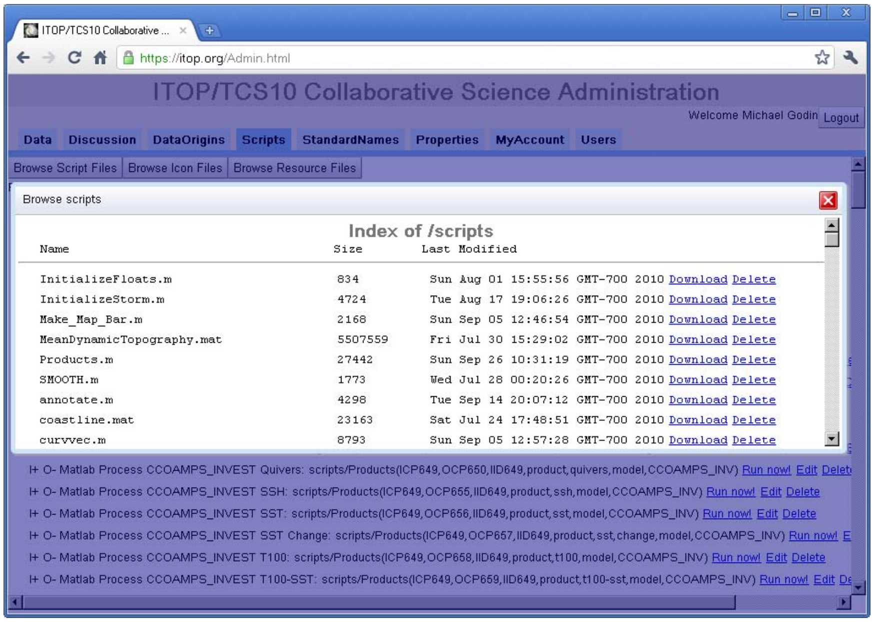

Figure 2, Graphical User Interface for managing product generation scripts that run automatically on the server when new data is received or at periodic intervals.

A critical capability identified in 2003 and 2006 as desirable in future field programs are tools that allows "power users" to identify and describe new data streams so they can be integrated into the system, and tools that monitor the progress of the system in real-time, ensuring that data products were being collected at the right time, and notifying users of any problems in the process. The approach for ITOP includes such capabilities (see Figure 3).

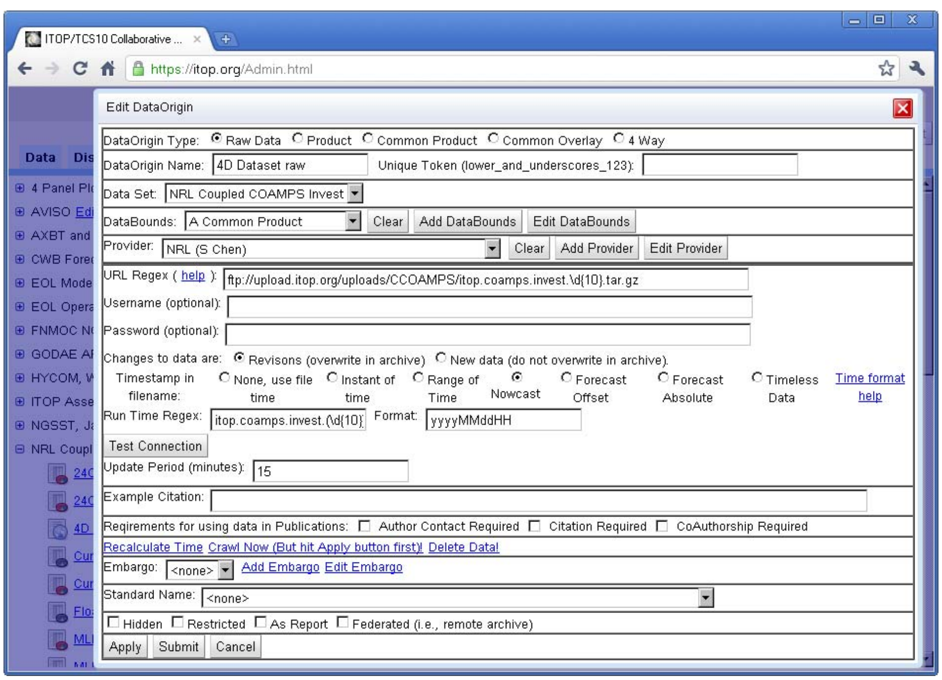

Figure 3, GUI for managing incoming data streams.

Another need expressed by ITOP researchers in the Pilot Experiment and Virtual Experiment was the desire to view geo-referenced products in Google Earth. For products generated by itop.org, which were generated with a projection that was compatible with Google Earth, this was easy. However, such was not the case with many externally-generated products, including those collected by EOL. For these products, a re-projector was developed which allowed the non-conforming products to be viewed along with conforming products in Google Earth (see Figure 4, Figure 5, and Figure 6).

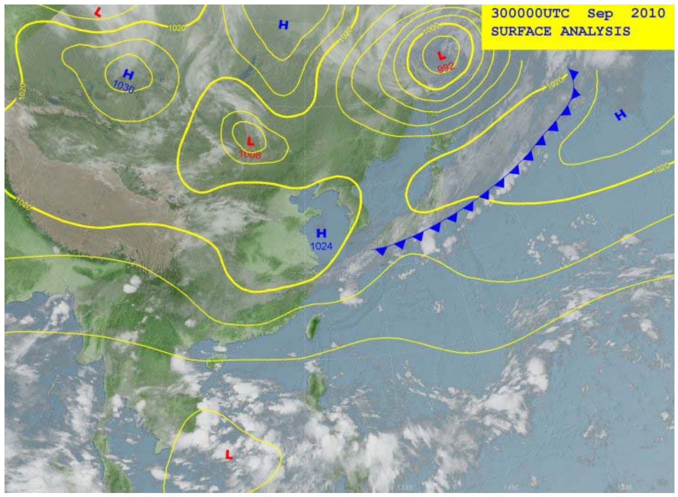

Figure 4, Original image obtained from Taiwan's Central Weather Bureau, in Lambert Conformal Conic projection

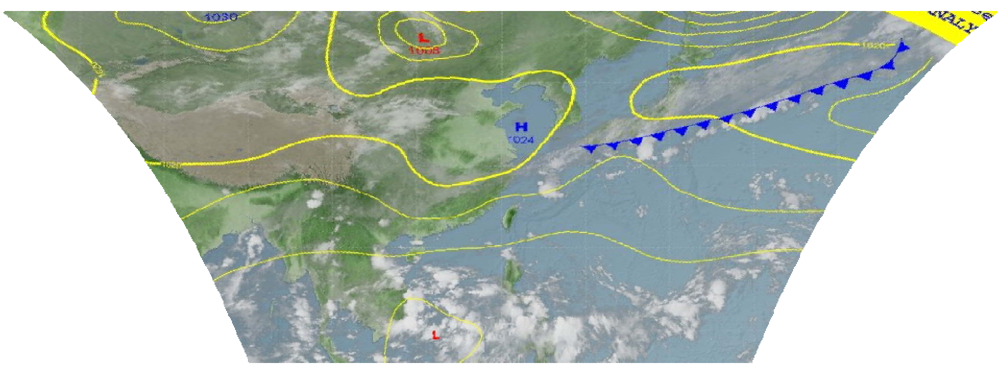

Figure 5, Image re-projected into Plate Careé projection for Google Earth

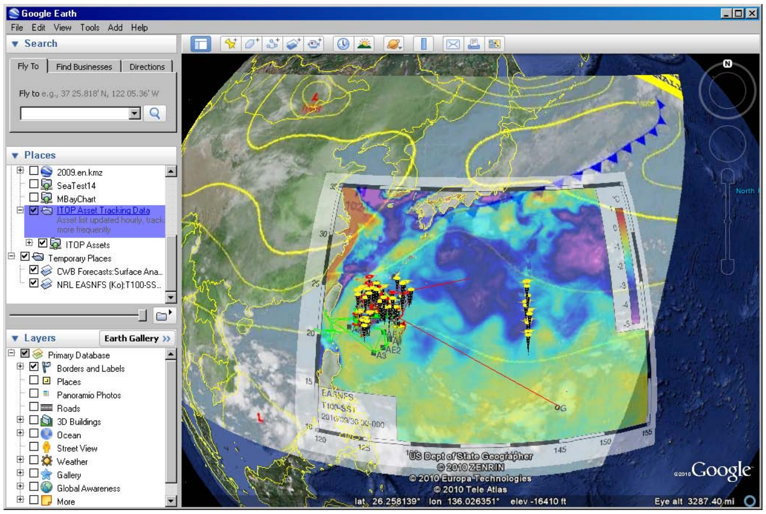

Figure 6, Image displayed in Google Earth along with other geo-referenced products.

RESULTS

As of September 30 2010, 522 collaborative discussion items have been posted to itop.org on 126 different topics. Additionally, 287,000 data files have been indexed by itop.org. Of these, 11,500 files (204 gigabytes) are downloaded raw data files, 36,000 files (3.4 gigabyte) are products generated from the raw data, 4,000 files (1 gigabyte) are products downloaded from external sources, and 216,000 files (33.5 gigabytes) are maintained on UCAR's EOL catalog and indexed on itop.org.

The implementation of CoSci web application runs entirely on compute cloud infrastructure, meaning that no hardware investment has been necessary, and no hardware maintenance will be required in the future. A small monthly bill covers the cost of compute cycles on the virtual server hosting the application, the cost of data transfer to and from the server, and the cost of data storage. Most importantly, running CoSci in the "cloud" means that duplicating the system for another field program would take hours, rather than months that would be required to procure, install, and configure a physical server.

IMPACT/APPLICATIONS

The success so far with CoSci has shown that for future field programs and observatories, ONR will have the option of quickly and easily deploying CoSci portals.

A CoSci portal has already been launched for the 2010 NRL Biospace / MBARI BloomEx experiment occurring in October 2010 (see Figure 7). This deployment is sponsored by the Packard Foundation, which will support further development of CoSci to meet the additional needs of Biospace/BloomEx researchers, making CoSci a more attractive tool for adoption by future field programs.

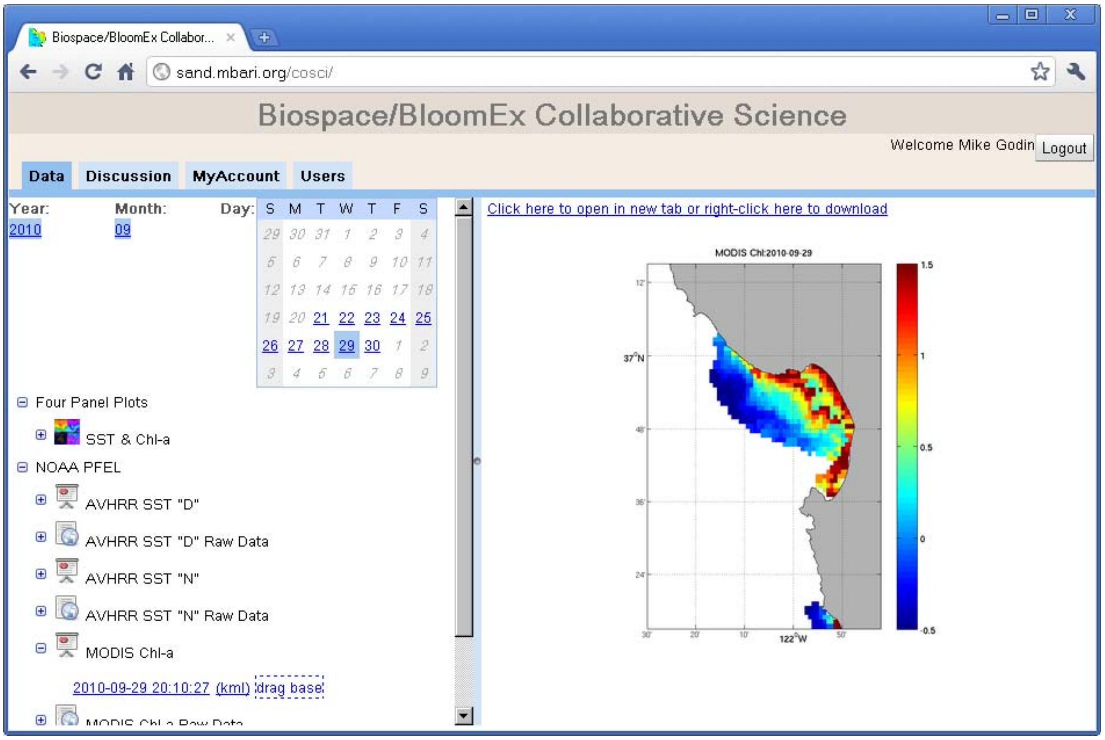

Figure 7, CoSci portal being used for 2010 NRL Biospace / MBARI BloomEx field program.

REFERENCES

[i] Huergo, J., "ONR Supports Monterey Bay 2006 Experiment", *ONR Media*, Aug. 23 2006 <http://www.onr.navy.mil/media/article.asp?ID=112>.

[ii] Bellingham J. G. and Godin, M., "Exploring Ocean Data," *ACM Sigmod*, Jim Gray Tribute Issue, 37(2) 78-82, 2008.

[iii] Godin, M.A.; Bellingham, J.G.; Rajan, K.; Leonard, N.; Yi Chao, "A Collaborative Portal for Ocean Observatories," *OCEANS 2006* , vol., no., pp.1-5, Sept. 2006

[iv] Murray, Don, "IDV at 5: New features and future plans," 25th Conference on International Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology, Jan 10-15, 2009

[v] Godin, M.A.; Bellingham, J.G., "Data Exploration for Multidisciplinary Research," *Oceans 2007*, vol., no., pp.1-4, Sept. 29 2007-Oct. 4 2007

[vi] Stefan Hepper, *What's new in the Java Portlet Specification V2.0 (JSR 286)*, http://www.ibm.com/developerworks/websphere/library/techarticles/0803_hepper/0803_hepper.html, 2008